

The Four Basic Types of Evaluation: Clinical Reviews, Clinical Trials, Program Reviews, and Program Trials

J. H. ABRAMSON, BSc, MB, BCh

IN RECENT YEARS so much has been written about the evaluation of health care, and so many models and techniques have been described (1-7), that it is difficult to see the forest for the trees, let alone find the best path through it. This confusion, although not always conscious and seldom expressed, is often made obvious by the poor design or unhelpful findings of evaluative studies. Therefore, it is worthwhile to try to simplify matters by distinguishing between the various basic situations in which evaluation may be performed and to see how tasks and methods vary in these situations.

The Basic Types of Evaluation

Four basic types of evaluation can be usefully identified; each is appropriate in a distinctive situation. The differentiation of these situations depends on the answers to the following two questions:

Why is the evaluation being done? Is it primarily motivated by concern for a specific person or number of persons for whom care is provided, or is it motivated by a wish to generate knowledge of more general applicability, for example, concerning the value of a new form of treatment or health program?

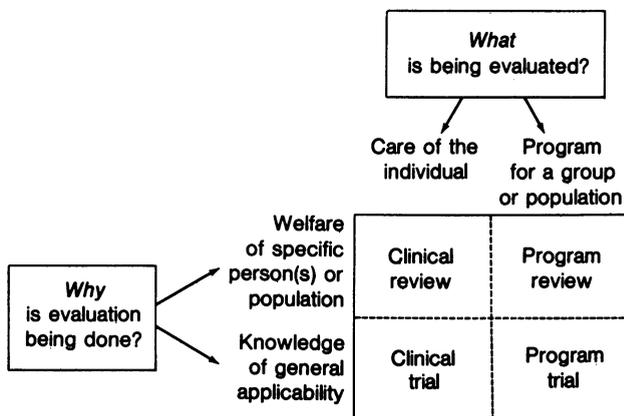
What is being evaluated? Is the care of the individual or the care directed at a group or population being evaluated? The care of the individual can be construed here as including the care of the individual family, if the family is the unit of care. Care directed at a group or population is referred to as a "program."

This term is used here in a broad connotation, to signify not only a program with well-defined goals (case-finding, the control of hypertension, smoking cessation, or fluoridation, for example) but also a total service (a national or regional health service, a health center, a hospital, or a general practice), a part or aspect of such a service, the work of a specific category of health worker, or even the work of a specific practitioner.

The basic types of evaluation are shown in figure 1. They are separated by broken lines in the diagram to stress that hybrid situations—in which a combination of different types of evaluation may be called for—commonly occur. In a routine clinical service, for example, both the care of particular persons and the service as a whole usually will be evaluated. In testing a new form of treatment, the aim may be to appraise its value for the individual patient and to evaluate the program whereby it is made available to the public.

The *clinical review* is the most common type of evaluation in the health field. It is a review of the care received by a specific patient or family, performed because of interest in the welfare of this patient or family and to provide a basis for decisions on whether to con-

Figure 1. The four basic types of evaluation



Dr. Abramson is a professor of social medicine at the Hebrew University-Hadassah Medical School and an Established Investigator of the Chief Scientist's Bureau, Ministry of Health, Israel. Tearsheet requests to Dr. J.H. Abramson, Department of Social Medicine, Hebrew University-Hadassah Medical School, P.O. Box 1172, Jerusalem, Israel.

tinue, change, or stop treatment. The term clinical review is used here to refer only to this form of appraisal, which is a basic element of the clinical process. The care provided to individual patients may also be appraised (usually more formally) for other purposes, for example, as a technique for evaluating a health service.

A *clinical trial* is an appraisal of the worth of a specific type of care given to the person. The term is used broadly here to include experimental and non-experimental evaluations of therapeutic, preventive, and other procedures, including diagnostic and educational ones. A trial sets out to obtain generalizable knowledge that can be applied in other settings; it must yield conclusions that are sufficiently substantiated to be generally convincing.

A *program trial* similarly aims to yield well-grounded, convincing, and generalizable conclusions, this time concerning the value of a specific type of health program. The program may be an established one or one set up as a test or demonstration.

A *program review* appraises a specific program, rather than a type of program. It is concerned with a particular program that operates in a defined setting, and it is motivated by concern for the specific patients or population the activities are designed to help. An essential feature is that the findings should be helpful to those who make decisions about the program. The evaluation can therefore be conducted within the framework of the assumptions accepted by these decision makers, for example, the assumption that the performance of certain procedures will have beneficial effects. These assumptions, on which the program is based, are not necessarily questioned or tested. The evaluation results are "presumptive" rather than "definitive" (8), and they can be useful to those responsible for the service without necessarily being found convincing by those who doubt the validity of the assumptions on which the program is based.

The distinction between what I call program reviews and program trials, which differ in their need to yield conclusions that are sufficiently well-grounded to be generally accepted, is particularly important. Failure to draw this distinction is a frequent cause of wasteful or unhelpful evaluative studies. The analogies drawn between program reviews and clinical reviews, and between program trials and clinical trials, may help to clarify this distinction.

Basic Questions

The four types of evaluation differ in the questions they pose and in the methods used to answer them. The difference in questions lies not so much in the nature of these questions as in their relative importance.

The following are the basic questions commonly posed in evaluative studies, whether as separate issues or as components of global appraisals. As used here, "care" refers to whatever action is being evaluated, including screening, casefinding and other diagnostic activities, and nonpersonal health programs.

Requisiteness: *To what extent is care needed?* Appraisal of the requisiteness or "appropriateness" (9) of care is not only based on information on the size and impact of the problem or set of problems that necessitates care, but it is also influenced by other considerations, such as the nature and severity of competing problems.

Quality: *How satisfactory is the outcome—attainment of desirable effects (effectiveness) and absence of undesirable effects (harmlessness)?* A full appraisal of outcome depends on the balance between desirable effects (planned or unplanned) and undesirable ones (anticipated or unanticipated). The ultimate criterion is the extent to which the underlying problem is alleviated or prevented without producing or aggravating other problems.

How satisfactory is the performance of activities—by the providers of care and by the recipients of care (compliance, use of services, and community participation)? The activities of the providers and recipients of care are elements of the "process," as opposed to the outcome of care. The appraisal may be concerned with how these activities are performed and with quantitative aspects. It may extend to studies of knowledge and attitudes that may influence overt behavior and to studies of relationships and communication between providers and recipients of care.

How satisfactory are facilities and settings? Facilities and settings—equipment, personnel, and organizational and fiscal arrangements—may be seen as further elements of the process of care (10) or as the structure underlying this process (4).

Efficiency: *How efficiently are resources used?* Efficiency is a measure of the cost in resources that is incurred in achieving results: It is determined by the balance between input (in time, manpower, and equipment, or their monetary equivalent) and output. The input may be compared with measures of effectiveness (cost-effectiveness analysis) or with the monetary equivalent of the output. The "systems model" approach goes further and aspires to an appraisal of the extent to which there is an optimal allocation of resources within an organization which provides care, taking into account its multiple functions over and above the achievement of specific care objectives (11).

Satisfaction: *How satisfied are the people concerned?* Satisfaction usually refers to the attitudes of patients and the public.

Differential value: *How do the preceding features differ for different categories of patients or groups of people or in different circumstances?* Differential value includes the appraisal of differences between groups of people or categories of patients in their need for care, in the provision of care, and in favorable and adverse outcomes. Equality of service may be one touchstone in evaluating a program (10, 12).

The arrangement of items in this scheme is, of course, somewhat arbitrary. "Compliance" and "satisfaction," for example, are sometimes treated as types of outcome, and "differential value" may be considered under each of the other items rather than appearing separately.

Clinical Reviews

The key components of a clinical review are appraisals of the requisiteness and outcome of care. If the clinician judges that treatment is no longer needed, further evaluation is superfluous. If the need for treatment persists, effectiveness and undesired effects must be considered. Because of the importance of relationships with the patient and his or her family, emphasis may also be given to satisfaction with care. Other questions are usually subsidiary; if, for example, the outcome has been unsatisfactory, an explanation may be sought in poor compliance.

Clinical review is highly subjective. It is usually performed by the clinician, who may find it difficult to be impartial. It is often based on incomplete data, if only because of the need for rapid decisions. Full and direct measures of the effects of treatment are seldom available. Also, it is difficult to be sure that these effects are due to the treatment, unless they are highly specific or there is convincing evidence of dose-effect or time-effect relationships. The criteria used for appraising effectiveness are seldom specifically formulated and may differ among clinicians. Disagreement is frequent when different physicians appraise the care received by the same patients (13) or when physicians' appraisals are compared with those made by patients (14). Despite these shortcomings, the clinical review remains an indispensable tool in clinical work.

Clinical Trials

The key issues in clinical trials are usually effectiveness (efficacy) and harmlessness, in general and among patients with different characteristics. Attention may also

be focused on efficiency; for example, by a comparison of the costs of different ways to achieve a similar benefit. Appraisals of performance, compliance, satisfaction, and facilities and settings are usually seen as subsidiary, serving only to explain why the outcome is or is not satisfactory. However, these appraisals may be central features in trials that aim to evaluate the feasibility or acceptability of a procedure. The requisiteness of care is a precondition for the performance of a clinical trial, rather than a question posed in the trial.

If a clinical trial is to yield convincing conclusions, it must be designed and conducted with meticulous attention to detail; rigorous attention to accurate measurement and the avoidance of bias are required. Unless the effects of care are marked and specific, a convincing demonstration that they can be ascribed to the procedures being tested—rather than caused by extraneous "confounding" factors—requires special precautions, such as the use of controls and randomization. The demonstration of this cause-effect relationship is usually the prime aim of a clinical trial.

The outcomes used as criteria of effectiveness in clinical trials are usually health characteristics, or changes in health status, that are regarded as clearly desirable either in their own right or as proxy indicators or predictors of favorable consequences.

Program Trials

The questions posed in program trials are similar to those asked in clinical trials. The central issue is usually the causal relationship between care and outcomes. Attention may also be focused on economic efficiency. Appraisals of performance, compliance, satisfaction, facilities, and settings usually are specifically aimed at explaining effectiveness or its lack, except in studies that focus on feasibility or acceptability.

Effectiveness can be convincingly appraised only if the outcomes used as criteria are clearly desirable, either as end results in their own right or as steppingstones to such end results. In instances where there is no doubt of the benefits to be expected from an activity (for example, immunization), a measure of its performance may be used as a criterion of effectiveness.

As in clinical trials, the need to produce well-founded conclusions imposes considerable demands, which may be difficult to meet. Rigorous attention must be paid to accuracy and objectivity in measurement. A reasonably convincing demonstration of a cause-effect relationship usually necessitates the measurement of numerous possible confounding factors. Randomization is seldom feasible, but recourse to quasi-experimental methods is usually possible (15); such methods may require control groups or populations and serial measurements.

In appraisals of effectiveness, the goal attainment model (16) is often used. With this model, the appraisal is based on the extent to which the predetermined goals of the program are achieved. These goals are the previously selected desirable outcomes of the program (fig. 2). They may be "subgoals" (intermediate outcomes—for example, O_1 in figure 2), the "program goal" (the last outcome not followed by further program activities— O_2 in figure 2) or "ultimate goals" (O_3). This method requires the prior explicit formulation of goals in clear and measurable terms, a requirement usually met without much difficulty in program trials that have been planned as such. The use of this model—which does not usually allow for undesirable outcomes—presents problems if goals are not readily amenable to measurement, if they cover only part of the beneficial outcomes, and, of course, if they have not been fixed in advance and cannot be satisfactorily formulated in retrospect.

Whether or not this model is used, effectiveness can be convincingly appraised only if the outcomes used as criteria are clearly desirable, either as end results in their own right or as steppingstones to such end results. In a program trial, an intermediate outcome (O_1 in figure 2) is a convincing criterion only if the cause-effect link between it and a desirable end result is not open to question.

Program Reviews

The key questions in a program review are usually quality or efficiency, or both. Attention may also be focused on requisiteness, especially in a long-established program, on the public's satisfaction and on differences between population groups or categories of patients, especially in their need for care, the extent and use of services, and coverage.

The methods used are largely determined by three

characteristics of this type of evaluation, which stem from the fact that its main purpose—analogueous to that of a clinical review—is to provide a basis for decisions concerning a specific service or program in a specific setting. First, the validity of the assumptions underlying the program—for example, that certain activities will have desirable effects—may be taken for granted. This greatly simplifies the evaluation.

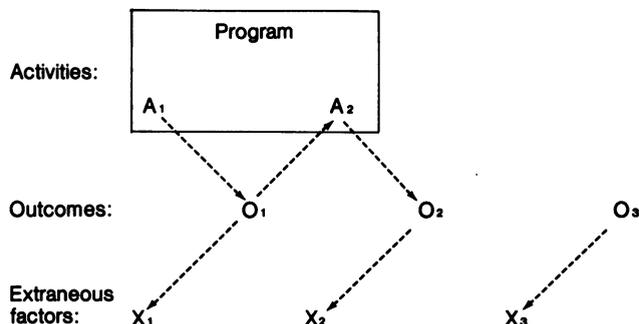
Second, for maximum value the evaluation must be rapid and, if possible, ongoing. Changes in circumstances, personnel, and policy result in frequent alterations in procedures, and there may be little practical benefit in evaluating a program as it used to be some years previously. If appraisal is rapid, it can give early warning of inadequacies and provide an up-to-date factual basis for decisions. Methods that require long-term followup, such as the measurement of remote end results, may fail to meet this need.

Third, a program review is carried out in a service-oriented setting—this is the review's *raison d'être*. Evaluation may not be seen as a central aim of the service, and little time or resources may be available for special information-collecting procedures. This situation may prevent a serious examination of cause-effect relationships, which may require studies of controls, measurements of numerous possible confounding factors, serial observations, and special precautions to enhance accuracy and reduce bias. Difficulties in obtaining information about those members of the target population with whom there is no routine contact, or with whom such contact has ceased, may also impede rigorous studies of requisiteness, outcome, compliance, and public satisfaction.

In these circumstances, measures of the "process" of care become very important. If it is accepted that the performance of certain activities, compliance with certain instructions, and the provision of certain facilities and equipment are desirable—although this is a matter of opinion rather than of established fact—then measures of these aspects are acceptable criteria of quality. Moreover, such data can usually be obtained relatively easily and rapidly, often as a byproduct of the routine work of the program, that is, by inbuilt monitoring procedures.

In a program with an explicitly formulated plan, activities can be appraised in terms of the extent to which they are performed as planned. If there is no record of what activities were planned—as is often the case in a clinical service—arbitrary standards may be used. In recent years much attention has been paid to methods of developing explicit criteria for this purpose, based on the opinions of experts (normative standards) or the practice of professionals (empirical standards). "Medi-

Figure 2. Chain of activities and outcomes



Activity A₁ (for example, screening for hypertension) leads to intermediate outcome O₁ (hypertensives in the population have been identified). Then activity A₂ (the treatment of hypertension) leads to outcome O₂ (a decreased prevalence of raised pressures in the population). This leads to outcome O₃ (a decrease in mortality from complications of hypertension). Extrinsic factors X₁, X₂, and X₃ may contribute to these outcomes. A program may comprise a number of parallel, branched, and interconnecting chains of this sort.

cal audit" and similar techniques for evaluating clinical care ("peer review," "self-audit," "medical care evaluation studies," or "nursing audit") are largely based on appraisals of performance and other components of the process of care (4-6, 17-20). The data come from routine records, from specially modified or designed records, or from special investigations, including direct observations of practitioners at work. The review may cover all patients cared for, a representative sample, or defined categories, such as patients with selected "indicator" conditions or those with poor outcomes.

Audit techniques are primarily applicable in evaluative reviews of clinical services. They are based on the assumption, seldom supported by convincing evidence (21), that the performance of certain procedures is likely to benefit patients. Important advantages of audit techniques are their educational value for personnel and the ease with which evaluation results can be translated into practical recommendations.

This emphasis on measures of process does not mean that measures of outcome have no place in a program review. On the contrary, information on outcomes may be valuable even without rigorous evidence that outcomes are actually consequences of the program. It is usually believed that the probability that outcomes are at least partly due to the program is strong enough to form a basis for decisions, or at the very least to indicate whether there is a need for more detailed evaluative study. Intermediate outcomes are usually the easiest to measure, but if data can be obtained on relevant end results, such as mortality rates, case fatality rates, or changes in the health status of patients or the population, such information is often especially helpful. If the program has predetermined goals, information on their accomplishment is, or course, particularly meaningful.

Much attention has recently been paid to measures of outcome in appraising the quality of clinical care (5,6,17). The outcomes measured include end results—"changes in the patient as a person or in the attributes of the disease or condition" (22)—and intermediate outcomes, such as the establishment of correct diagnoses or changes in the patient's health behavior. Although it is difficult to know to what extent outcomes can be attributed to medical care, this does not vitiate the usefulness of such measures in a program review. If patients do well, there is at least no cause for concern. Even a subjective appraisal of outcomes may be found helpful by those responsible for a program, though it may carry little conviction for others. An example is a review of a program for the housebound, evaluated on the basis of clinicians' appraisals of the extent of achievement of the rehabilitation goals that they had previously defined for each patient.

In a program review the appraisal of economic efficiency, like that of quality, has special features. Although detailed studies of inputs may be undertaken, emphasis is often put on simple observations that can be used as a basis for decisions aimed at enhancing efficiency. Such observations relate especially to evidence of wasteful operation—the avoidable use of expensive or ineffective drugs, overstaffing, delays, the underuse of expensive equipment, superfluous activities, unduly long institutional care, and the like. These observations may be appraised in terms of implicit or explicit standards. If cost-effectiveness studies are undertaken, cost is usually balanced against estimates or subjective appraisals of effectiveness, or against the performance of assumedly beneficial activities.

Discussion

Like other value judgments, the evaluation of health care can never be wholly objective. Objectivity can be enhanced, however, and its limits known and shown, if criteria are explicitly formulated and if careful attention is paid to the collection of accurate and pertinent information. The methods used should be suited to the needs and realities of the situation. It is wasteful and may be self-defeating to use demanding techniques in a situation where basic assumptions are not in question and no rigorous proofs are required. On the other hand, an evaluation may be inconclusive if insufficiently rigorous methods are used in a situation where well-substantiated and generally convincing conclusions are sought.

It has been suggested that detailed evaluations should be done only in selected situations, in such a way that their results are transferable to other settings, while "evaluation in the usual service program or project should be confined to a quality control type of evaluation based on process or intermediate goals. . . . The service program should not be undertaken unless the decision makers are willing to accept (either on the basis of previous evidence or faith) the premise that a properly conducted program of that type does ultimately have a beneficial effect on health" (23).

Sometimes, however, basic assumptions of this sort are brought into question in evaluations of established programs or services. The value of care procedures may vary in different circumstances, and "one can never be certain that a program that works in one situation will work in another" (1). Moreover, there may be a wish to appraise innovative features. Under these circumstances, there is a need for the demanding methods appropriate to a program trial, often in association with a less rigorous review of other aspects of the program. In principle, the performance of a program trial within the setting

of an established program presents a challenge which should be met when possible. In practice, however, difficulties usually abound, and such evaluations should not be undertaken lightly. They are likely to be successful only in exceptional circumstances, since evaluation and service frequently make competing demands, and the requirements for a well-substantiated evaluation of effectiveness may be difficult to meet in "the turbulent setting of the action program" (3).

A verdict on evaluation in the health field today might read "Too little and too much." Too little—whether because of lack of motivation, lack of forethought, lack of know-how, or for other reasons—with the result that procedures of unproved and often dubious worth remain in common use and decisions continue to be made on insufficient factual grounds. And too much, in the sense that evaluation is frequently so unduly detailed or rigorous that its performance becomes uneconomical, its results are unhelpful, and evaluation as a whole falls into disrepute. A fuller awareness of the different tasks of evaluation and the appropriate methods in different situations may in time help to change this verdict.

References

- Suchman, E. A.: *Evaluative research: principles and practice in public service and social action programs*. Russell Sage Foundation, New York, 1967.
- Schulberg, H. C., Sheldon, A., and Baker, F., editors: *Program evaluation in the health fields*. Behavioral Publications, New York, 1969.
- Weiss, C. H.: *Evaluation research: methods of assessing program effectiveness*. Prentice-Hall, Englewood Cliffs, N.J., 1972.
- Donabedian, A.: Evaluating the quality of medical care. *Milbank Mem Fund Q* 44: 166–203 (1966).
- Donabedian, A.: Promoting quality through evaluating the process of patient care. *Med Care* 6: 181–202 (1968).
- Donabedian, A.: *A guide to medical care administration*. Vol. II: *Medical care appraisal—quality and utilization*. American Public Health Association, Washington, D.C., 1969.
- Costanzo, G. A., and Vertinsky, I.: Measuring the quality of health care: a decision-oriented typology. *Med Care* 13: 417–431 (1975).
- Stanley, D. T.: Excellence in the public service: How do you really know? *Public Admin Rev* 24: 170–174 (1964).
- Kane, R. L., Henson, R., and Deniston, O. L.: *Program evaluation: Is it worth it?* In *The challenges of community medicine*, edited by R. L. Kane. Springer Publishing Co., Inc., New York, 1974, pp. 213–233.
- Doll, R.: Surveillance and monitoring. *Int J Epidemiol* 3: 305–314 (1974).
- Etzioni, A.: The approaches to organizational analysis: A critique and a suggestion. *Admin Sci Q* 5: 257–278 (1960).
- Cochrane, A. L.: *Effectiveness and efficiency*. Nuffield Provincial Hospitals Trust, London, 1972.
- Koran, L. M.: The reliability of clinical methods, data and judgments. *N Engl J Med* 293: 642–646, 695–701 (1975).
- Cay, E. L., et al.: Patient's assessment of the result of surgery for peptic ulcer. *Lancet* 1: 29–31 (1975).
- Campbell, D. T.: Factors relevant to the validity of experiments in social settings. *Psychol Bull* 54: 297–312 (1957).
- Deniston, O. L., Rosenstock, I. M., and Getting, V. A.: Evaluation of program effectiveness. *Public Health Rep* 83: 323–335 (1968).
- Barro, A. R.: Survey and evaluation of approaches to physician performance measurement. *J Med Educ* 48: 1047–1093 (1973).
- Blum, H. L.: Evaluating health care. *Med Care* 12: 999–1011 (1974).
- Brook, R. H.: Quality of care assessment: A comparison of five methods of peer review. U.S. Department of Health, Education, and Welfare, Public Health Service, Health Resources Administration, Rockville, Md., 1973.
- Brook, R. H.: Critical issues in the assessment of quality of care. In *The challenges of community medicine*, edited by R. L. Kane. Springer Publishing Co., Inc., New York, 1974, pp. 183–211.
- Donabedian, A.: Measuring and evaluating hospital and medical care. *Bull NY Acad Med* 52: 51–59 (1976).
- Sanazaro, P. J., and Williamson, J. W.: End result of patient care: A provisional classification based on reports by internists. *Med Care* 6: 123–130 (1968).
- Thorner, R. M.: Health program evaluation in relation to health programing. *HMSHA Health Rep* 86: 525–532 (1971).

SYNOPSIS

ABRAMSON, J. H. (Hebrew University-Hadassah Medical School, Jerusalem): *The four basic types of evaluation: clinical reviews, clinical trials, program reviews, and program trials*. *Public Health Reports*, Vol. 94, May–June 1979, pp. 210–215.

Four basic types of evaluation, each appropriate in a distinctive situation, are the clinical review and the clinical trial, which are con-

cerned with the care of the individual patient, and the program review and program trial, which deal with programs or services directed at groups or populations. Evaluative reviews are primarily motivated by concern with the welfare of the specific population served, and they appraise specific activities in specific settings as a basis for decisions concerning these activities. Clinical and program

trials aim to generate knowledge of more general applicability, especially concerning causal relationships between care and outcomes. The types of evaluation differ in the questions they pose and in the methods used to answer them. Failure to draw a distinction between program reviews and program trials is a frequent cause of wasteful or unhelpful evaluative studies.